

## **Identificação e caracterização *in silico* de retrotransposons Ty3-Gypsy no genoma de *Porphyra umbilicalis* (Bangiales, Rhodophyta)**

Renan S. OLIVEIRA <sup>1</sup>, Daniela MILSTEIN <sup>1</sup>

<sup>1</sup> Instituto do Mar, Universidade Federal de São Paulo, UNIFESP – Baixada Santista, SP

### **RESUMO**

**Introdução:** Desde que descobertos por Barbara McClintock, em milho, os elementos de transposição têm sido amplamente estudados em animais, fungos e plantas. No entanto, poucos trabalhos com objetivo de identificar e classificar esses elementos em algas têm sido realizados. Deste modo, o objetivo deste estudo foi realizar uma análise *in silico* da diversidade de retrotransposons da superfamília Ty3-Gypsy na macroalga *Porphyra umbilicalis*.

**Resultados:** Foram caracterizados pela primeira vez os LTR-RTs da superfamília Ty3-Gypsy no genoma de uma macroalga vermelha. Utilizando dados genômicos disponíveis publicamente, foi possível obter uma visão inicial da diversidade filogenética e estrutural desses elementos em Bangiales. Foram identificadas 10 novas famílias de cromovírus distribuídas em 6 novos clados.

**Conclusões:** Este estudo estabelece duas importantes questões ou hipóteses gerais. Primeiro, a espécie *P. umbilicalis* possui 10 novas famílias que foram descritas e compreendem diversos elementos diferentes de retrotransposons Ty3-Gypsy. Uma segunda hipótese importante deste estudo é que, assumindo que a transmissão horizontal de cromovírus tenha sido muito pouco frequente, a origem de pelo menos um clado de retrotransposons Ty3-Gypsy de *P. umbilicalis* pode ter sido anterior a divisão de Rhodophyta e plantas verdes.

**Palavras-chave:** cromovírus, macroalga, retrotransposons, Ty3-Gypsy.

## INTRODUÇÃO

Elementos de transposição (TEs, do inglês *Transposable elements*) são fragmentos de DNA que podem ser inseridos em novos locais cromossômicos, podendo muitas vezes fazer cópias de si mesmos no processo (Feschotte et al. 2002). Até o momento, e com poucas exceções, os TEs são descritos em todos os genomas eucarióticos já estudados (Feschotte et al. 2002).

Apesar de comporem grande parte do genoma de muitas espécies (Kidwell & Lisch 2001), inicialmente, estes elementos eram chamados de “DNA lixo”, visto a aparente falta de função no organismo hospedeiro (Ohno 1972). No entanto, com o avanço das técnicas moleculares e consequentemente, o crescente estudo de genomas de diversas espécies, observou-se que os TEs, além de serem grandes responsáveis pela expansão e evolução de genomas (Hawkins et al. 2009), também podem dar origem a novos genes (Volff 2006), influenciar na expressão gênica (Herpin et al. 2010; Rebollo et al. 2012) e no rearranjo cromossômico (Zhang et al. 2009).

Dentre os TEs, os retroelementos com longas repetições terminais (LTR-RTs), também chamados de retrotransposons ou retroelementos, são os elementos móveis capazes de gerar cópias de si próprio por intermédio de um RNA mensageiro (Finnegan 1992). Estruturalmente, esses elementos possuem em suas extremidades 5' e 3' longas terminações repetidas diretas (LTRs), que são delimitadas pelos sítios TSDs (do inglês, *Target site duplication*). O sítio TSD é formado pela reparação de cortes coesivos gerados durante a transposição do retroelemento, sendo assim um ponto de identificação da inserção do retrotransposon no genoma hospedeiro (Kazazian 2011). Entre as LTRs há dois genes responsáveis pelo ciclo de transposição do elemento: *gag* e *pol*. O gene *gag* codifica uma proteína similar ao capsídeo dos retrovírus e o gene *pol* codifica um complexo poliproteico que inclui os domínios protease (PR), transcriptase reversa (RT), RnaseH (RH) e integrase (INT) (Feschotte et al. 2002; Novikova, 2009).

Inferências filogenéticas baseadas nas sequências dos domínios RT, RH e INT, classificam os retrotransposons com LTRs em três superfamílias (Carareto et al. 2015), sendo que *Ty1-Copia* e *Ty3-Gypsy* são as duas únicas superfamílias que ocorrem em plantas (Paz et al. 2017). Estruturalmente, os elementos pertencentes a *Ty1-Copia* e *Ty3-Gypsy* diferenciam-se entre si pela ordem do arranjo dos domínios RT, RH e INT no gene *pol* (Eickbush & Jamburuthugoda 2008). A superfamília *Ty3-Gypsy* é amplamente distribuída pelos diversos filos de eucariotos (Novikova 2009), e pode ser subdividida nas linhagens cromovírus e Tat/Athila. (Neumann et al. 2019).

Os cromovírus apresentam, além das proteínas típicas de LTR-RTs, um cromodomínio (CHD) na região C-terminal da integrase. Os CHDs estão presentes em várias proteínas eucarióticas envolvidas na remodelação da cromatina e na regulação da expressão gênica (Koonin et al. 1995; Neumann et al. 2019). Nos cromovírus esse domínio está relacionado com integração de novas cópias do retroelemento em posições específicas do genoma, como por exemplo regiões de heterocromatina e centroméricas (Gao et al. 2008; Novikova 2009; Nunes et al. 2018). Atualmente, são conhecidos seis clados de cromovírus: Chlamyvir, que ocorre especificamente em alguns grupos de algas; Tcn1, grupo específico de fungos; Galadriel, presente em plantas sem sementes e plantas com sementes; e as linhagens típicas de plantas com sementes: CRM, Del/Tekay e Reina (Neumann et al. 2019).

Desde que descobertos por Barbara McClintock na década de 1940 (McClintock 1951), os TEs tem sido amplamente estudados em animais, fungos e plantas com sementes (Pray 2008). No entanto, trabalhos identificando esses elementos em algas ainda são escassos. Retrotransposons são reportados para algas verdes (Lindauer et al. 1993, Perez-Alegre et al. 2005), cryptophytas (Khan et al. 2007), diatomáceas (Kohany et al. 2006) e algas vermelhas (Peddigari et al. 2008). Os LTRs-RT identificados em algas vermelhas foram isolados de *Pyropia yezoensis*, e pertencem à superfamília de retroelementos *Ty1-Copia* (Peddigari et al. 2008).

O gênero *Pyropia* assim como o gênero *Porphyra* (ordem Bangiales, Rhodophyta), incluem espécies de grande importância econômica. Essas espécies, popularmente conhecidas como “nori”, apresentam alta quantidade de minerais, vitaminas e proteínas (Noda, 1993; MacArtain et al. 2007; Smith et al. 2010), sendo assim bastante apreciadas na culinária asiática onde seu valor de produção anual é estimado em ~1,3 bilhões de dólares (Blouin et al. 2011). Embora não seja a principal espécie comercializada, *Porphyra umbilicalis* tem demonstrado grande potencial econômico em sistemas de aquicultura multitróficos, devido sua alta capacidade de absorção de nitrato (Blouin et al. 2007). A importância econômica e evolutiva dessa espécie levou ao sequenciamento completo de seu genoma para o desenvolvimento de um sistema modelo de estudo (Gantt et al. 2010).

Nos últimos anos, uma grande quantidade de dados genômicos de diferentes linhagens eucarióticas foi gerado, permitindo assim a análise da diversidade, evolução e origem dos retrotransposons em Eucariotos (Kordis 2005, Carareto et al. 2015, Paz et al. 2017). O genoma nuclear de *Porphyra umbilicalis* (87,7 Mbp), é composto por uma grande quantidade de elementos repetitivos (43,9%), onde os elementos de transposição são um dos mais representativos (Brawley et al. 2017).

Deste modo, o objetivo deste estudo foi aproveitar os recursos genômicos disponíveis de *P. umbilicalis*, em bancos de dados públicos, para realizar uma análise *in silico* da diversidade de retrotransposons da superfamília *Ty3-Gypsy* nessa espécie.

## **MATERIAIS E MÉTODOS**

### **Identificação de retroelementos LTR**

O *draft* do genoma v. de *Porphyra umbilicalis* foi obtido no banco de dados GenBank (Nº de acesso: GCA\_002049455.2). Os softwares LTR\_finder (Xu & Hao 2007) e LTRharvest (Ellinghaus, Kurtz & Willhoeft 2008) foram utilizados para uma triagem inicial de sequências que apresentassem LTRs com comprimento entre 100-7000 pares de base (pb), e região interna entre 1000-20000 pb. As sequências identificadas foram classificadas como possíveis retrotransposons. Para evitar falsas predições e também isolar sequências não redundantes de retrotransposons, uma segunda triagem foi realizada utilizando o software LTRretriever (Ou & Jiang 2018). As sequências classificadas pelo programa como *Gypsy* e *Unknown* foram selecionadas para análises posteriores.

### **Identificação de domínios protéicos e análises filogenéticas**

Para a identificação de possíveis retroelementos intactos foi realizada uma busca por domínios proteicos (GAG, RT, RH, INT e CHD) utilizando o software blastx (*E-value*: 1e-10), tendo como referência os bancos de dados GyDB (Llorens et al. 2010), REXDB (Neumann et al. 2019) e CDD-NCBI (Marchler-Bauer et al. 2016). As sequências que apresentaram *hits* com no mínimo três destes domínios foram retidas. As sequências dos domínios RT, RH e INT de cada retroelemento foram traduzidas em aminoácidos utilizando a ferramenta *transeq*, presente no pacote EMBOSS (Rice, Longden & Bleasby 2000). O alinhamento dos domínios de cada elemento foi realizado junto a sequências de linhagens de LTR-retrotransposons já conhecidas, utilizando a ferramenta MUSCLE (Edgar 2004) com parâmetros padrão, e verificados manualmente no programa Ugene (Okonechnikov 2012). Posições alinhadas incorretamente foram corrigidas, seguindo a similaridade de aminoácidos proposta por Llorens et al. (2009), e regiões com baixa qualidade (*gaps* > 90%) foram removidas a fim de obter um alinhamento de melhor qualidade. As árvores filogenéticas foram inferidas utilizando o método de *Neighbor-joining* (NJ) (Saitou & Nei 1987) implementado no software ClustalW (Thompson, Higgins & Gibson 1994), a partir do alinhamento concatenado das regiões RT-RH-INT (593aa) das sequências identificadas neste trabalho, e sequências de referências para cada clado conhecido de cromovírus. Como grupo externo foram utilizadas

sequências da linhagem de retrotransposons Athila/Tat. Para avaliar a confiança da topologia da árvore apresentada, foi utilizado o teste de *bootstrap* (Felsenstein, 1985) com 100 réplicas. As árvores geradas foram visualizadas e editadas na plataforma iTOL (Letunic & Bork 2006).

## **RESULTADOS**

### **Identificação de LTR-Retrotransposon**

A partir do *draft* do genoma de *Porphyra umbilicalis* foram identificadas 423 sequências putativas de retrotransposons. Destas, 96 foram classificadas como *Ty1-Copia*, 176 como *Ty3-Gypsy* e 151 como desconhecidas (*unknown*). As identificações basearam-se em características estruturais comuns a todos os LTR-RTs de plantas: a presença e tamanho de LTR-5' e LTR-3', e TSDs delimitando as LTRs. Como as LTRs 5' e 3' são idênticas no momento da inserção de uma nova cópia do elemento no genoma, o nível de divergência causado pelas mutações adquiridas ao longo do tempo é proporcional à idade de inserção. Para identificar os elementos Gypsy com possível inserção recente, e consequentemente intactos, foram selecionados apenas aqueles que apresentaram no mínimo 90% de similaridade entre as LTRs 5' e 3'. O tamanho das sequências inteiras dos elementos identificados como Gypsy variou de 3597pb a 13439pb, e o das LTRs de 341pb a 901pb. A grande maioria dos elementos (96,6%) é flanqueada por TSDs de 5pb, e os demais são flanqueados por TSDs de 4pb. Foram encontradas 156 combinações para TSDs de 5pb, sugerindo assim, que a integração destes elementos em novos locais ocorre de modo aleatório.

### **Identificação de domínios protéicos**

A fim de definir os elementos com domínios proteicos adequados para identificação e análises filogenéticas, as sequências de *Porphyra* foram comparadas com o NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al. 2016), visando identificar os elementos que apresentassem domínios com resíduos de aminoácidos altamente conservados, sendo assim, elementos possivelmente intactos. Em todas as sequências classificadas como *Ty3-Gypsy* foram analisados os domínios GAG, PROT, RT, RH, INT e cromodomínio. Sequências de elementos que não apresentaram no mínimo três domínios, ou que mostravam domínios duplicados foram removidas, visto que poderiam ser elementos preditos incorretamente. Ao total, 163 sequências apresentaram os domínios proteicos RT, RH e INT conservados. Essas sequências foram identificadas em 10 famílias, de acordo com a classificação proposta por Wicker et al. (2007).

## **Análises filogenéticas e identificação de clados**

As análises filogenéticas das sequências de aminoácidos dos domínios RT, RH e INT agruparam as 10 famílias identificadas, em 6 clados com alto suporte (bootstrap >95%) (Fig. 1). Características estruturais como comprimento de LTRs e tipo de cromodomínio corroboraram o agrupamento dos elementos de *P. umbilicalis* nos seis clados distintos (Fig. 2 e 3). Para cada clado definido, como mostra a Figura 1, foi selecionado um elemento canônico, ou seja, aquele com maior quantidade de domínios conservados. Cada um destes elementos foi utilizado para descrever o clado ao qual pertence.

O clado A, que inclui a família 1, é formado por 12 retroelementos. O elemento estruturalmente melhor conservado é encontrado no scaffold 23, e apresenta 12370 pb com 98% de identidade entre as LTRs de 649 pb. Possui duas ORFs (do Inglês, *Open Reading Frame*) de 423 e 1320 aminoácidos em frames diferentes. Foi o único clado que não apresentou cromodomínio em nenhuma de suas sequências.

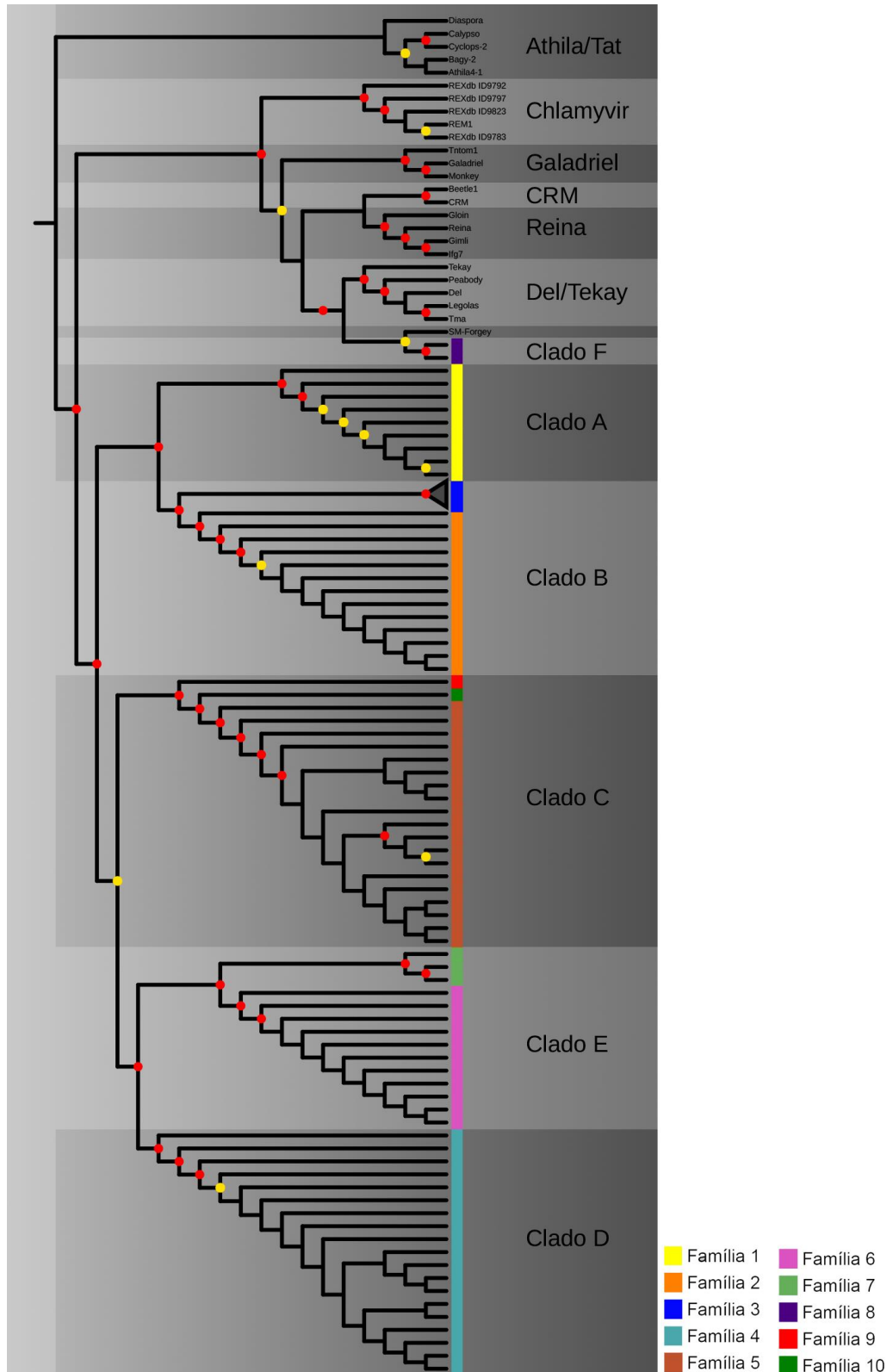
O clado B é proximamente relacionado ao clado A. Com 105 sequências é o clado com maior número de elementos presente no genoma de *P. umbilicalis*, sendo composto pelas famílias 2 e 3. A sequência canônica é encontrada no scaffold 1602 e tem 9135 pb, com LTRs de 395 pb que são 99% idênticas. Possui duas ORFs de 548 e 1445 aminoácidos em frames diferentes. Apesar de não ser possível identificar o domínio *gag*, no final da primeira ORF é encontrado o motivo CX2CX4HX4C, que geralmente ocorre ao final da *gag*.

Composto por 23 sequências, o clado C é formado pelas famílias 5, 9 e 10. Os elementos destas famílias, em sua grande parte (87%), não apresentam, ou apresentam um cromodomínio muito fragmentado. Para a cópia canônica foi selecionado o elemento localizado no scaffold 12, que possui 8510 pb com LTRs de 493 pb que são 98% idênticas. É possível identificar nesta sequência duas ORFs, em frames distintos, com 501 e 1209 aminoácidos, respectivamente. Em apenas poucos elementos deste clado (13%) foi possível observar ao final da segunda ORF, resíduos de aminoácidos típicos de cromodomínios bem conservados.

Os clados D e E são proximamente relacionadas e são representadas por 20 e 14 elementos, respectivamente. O clado D é formado pela família 4 e tem sua sequência canônica localizada no scaffold 47, com 10329 pb e LTRs de 768 pb que são 99% idênticas. Duas ORFs contíguas de 298 e 1304 aminoácidos são aparentes nessa sequência. No final da integrase, um cromodomínio é observado. O clado E inclui as famílias 6 e 7, e a sequência canônica, está localizada no scaffold 1626 com 9368 pb e LTRs de 748 pb que são 99%

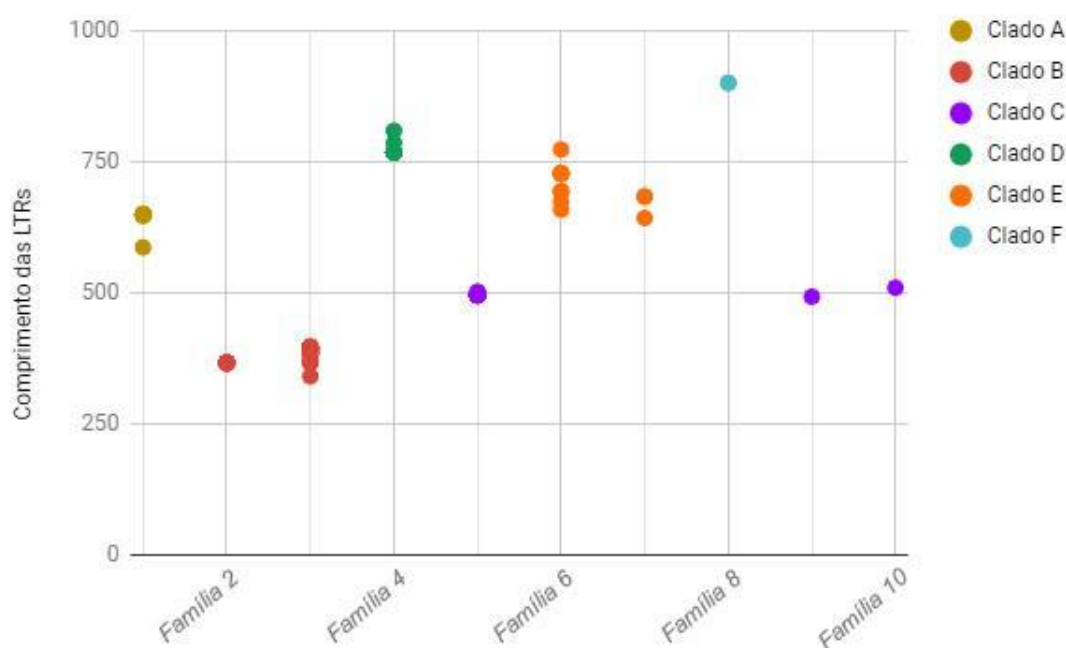
idênticas. Duas ORFs contíguas de 330 e 1350 aminoácidos são aparentes nessa sequência. No final da integrase, um cromodomínio é observado.

O clado F é o menor dos descritos no presente trabalho, é composto por apenas duas sequências, que formam a família 8, e tem o elemento melhor conservado localizado no scaffold 91. Apresenta 7202 pb e possui 99% de identidade entre as LTRs, que possuem 901 pb. Duas ORFs contíguas de 274 e 1094 aminoácidos são aparentes nessa sequência. É o único clado em que o domínio típico da proteína *gag* foi identificado.



**Fig. 1:** Dendograma inferido a partir de seqüências concatenadas de RT-RH-INT de todos os elementos Ty3-Gypsy encontrados em *P. umbilicalis* e seqüências de referência para cada clado de cromovírus, obtidas nos bancos de dados GyDB e RexDB. A árvore foi calculada usando o método de *Neighbor-joining* com 100 réplicas de *bootstrap*. Os círculos amarelos e vermelhos indicam *bootstrap* >50 e >70, respectivamente.





**Fig. 2:** Gráfico de dispersão indicando o tamanho característico das LTRs em linhagens clados distintos.

### Identificação de cromodomínios

Foi possível identificar 134 sequências com cromodomínio bem conservado na região C-terminal da integrase. Os cromodomínios podem ser divididos em dois tipo: 1. CHDs do tipo I apresentam três resíduos conservados, Y24, W45 e Y48 (posições 2, 28 e 31 do alinhamento, respectivamente); 2. CHDs do tipo II que na maioria das vezes possuem apenas o resíduo W45 conservado (Fig. 3).

Analisando o genoma de *P. umbilicalis*, foi possível identificar retroelementos LTR que apresentam CHDs do tipo I e do tipo II. Com base no alinhamento de CHDs (Fig. 3) foi possível observar que os clados E e F apresentam CHDs do tipo II. O clado E carece do primeiro resíduo de aminoácido Y24, mas mantém os aminoácidos W45 e Y48. Já o clado F, possui os resíduos Y24 e W45, mas carece do Y48. Os demais clados descritos no presente trabalho, exceto pelo clado A que não apresentou nenhum CHD intacto, apresentaram cromodomínio do tipo I. Os CHDs também apresentam um motivo característico chamado de *chromo-box*, cujas sequências conservadas dos aminoácidos estão representadas na Fig. 4.



esses elementos em algas têm sido realizados (Lindauer et al., 1993, Perez-Alegre et al., 2005, Khan et al., 2007, Kohany et al., 2006). No caso das algas vermelhas (Rhodophyta), esta situação é ainda mais crítica, pois os únicos elementos completos que se tem conhecimento, foram isolados da alga *Pyropia yezoensis* (como *Porphyra yezoensis*, Peddigari et al. 2008). Os LTRs-RT identificados em *Py. yezoensis* pertencem à superfamília de retroelementos *Ty1-Copia*. Além desses, também foi identificado na mesma espécie, por meio de análises de dados públicos, um fragmento de cromodomínio típico de cromovírus (Gorinsek, Gubensek & Kordis 2004).

O gênero *Pyropia* pertence a ordem Bangiales, linhagem que se diversificou a cerca de 1,5 bilhões de anos (Hedges 2002), e portanto, a descoberta de um fragmento de um cromodomínio revela o quão antiga é a origem dos cromovírus (Kordis 2005). De acordo com Gantt et al. (2010), o estudo dos TEs no genoma das Bangiales, aumentará a compreensão, ainda que limitada, acerca do processo evolutivo das algas vermelhas.

Neste trabalho, foram identificados e caracterizados pela primeira vez os LTR-RTs da superfamília *Ty3-Gypsy* no genoma de uma macroalga vermelha. Utilizando dados genômicos disponíveis publicamente, e com análises *in silico*, foi possível obter uma visão inicial da diversidade filogenética e estrutural desses elementos em Rhodophyta. Com o aumento do número de genomas sequenciados, os bancos de dados tem se tornados repositórios ricos em informações que permitem estudos dos elementos de transposição em diversos grupos de organismo (Llorens et 2009, Paz et al. 2017, Neumann et al 2019).

A topologia da árvore filogenética obtida neste trabalho, é similar a observada por Llorens et al. (2009), Puterova et al. (2017) e Neumann et al. (2019), onde os clados de cromovírus presente em plantas, Chlamyvir, Tcn1, Galadriel, CRM, Del/Tekay e Reina, apresentam o mesmo padrão de agrupamento. Dentre as dez novas famílias de retrotransposons identificadas no presente trabalho, de acordo com a classificação de Wicker et al. (2007), foi possível observar características estruturais similares entre algumas, como o tamanho das LTRs, que serviu de base para corroborar os seis novos clados de cromovírus identificados a partir da inferência filogenética.

Os clados A, B, C, D e E formaram um agrupamento monofilético, o que confirma a previsão de Kordis (2005) e Neumann et al. (2019) da possível existência de pelo menos um clado específico de cromovírus em Rhodophyta. O clado F é filogeneticamente mais distante dos demais clados encontrados em *P. umbilicalis*, e agrupa com o retroelemento SM-Fogey, descrito na licófito *Selaginella moellendorffii*. A cópia do retroelemento SM-Fogey descrita por Novikov et al. (2012) é estruturalmente muito diferente dos elementos do clado F. O

retroelemento SM-Fogey é relativamente menor, com tamanho de cerca de 5kb, e com LTRs de 233 pb. Além disso, não apresenta o cromodomínio na região C-terminal da integrase. SM-Fogey é considerado um elemento não classificado, visto que sua relação filogenética com os demais clados de cromovírus, ainda é incerta (Novikov et al. 2012, Neumann et al. 2019).

Clados de cromovírus específicos de plantas sem sementes já foram relatados na literatura (Novikova et al. 2008, Novikova, Smyshlyaev & Blinov 2010). No entanto, devido a disponibilidade de poucos genomas pertencentes a este grupo, pouco se sabe sobre a estrutura destes elementos e sua história evolutiva. Apesar das diferenças estruturais observada entre os elementos do Clado F e do elemento SM-Fogey, assim como, baixo suporte filogenético no ramo que agrupa ambos os elementos, os elementos do clado F estão mais próximos do elemento SM-Fogey, do que dos demais elementos de Rhodophyta. Isso abre a possibilidade de existirem mais clados que compreendam cromovírus de algas vermelhas e plantas que, devido estudos envolvendo estes dois grupos ainda não terem sido realizados, permanecem desconhecidos, embora a evolução dos retroelementos com cromodomínio em algas vermelha, tenha se dado de maneira independente dos cromovírus de plantas verdes (Llorens et al. 2009, Lee et al. 2018). posição incerta

Já foi observado que vários fatores podem influenciar a diversidade de retrotransposons em um genoma. Quais?

Espécies da ordem Bangiales apresentam em sua fase esporofítica uma ligação estrutural entre células adjacentes, conhecidas como *pit connection* (Lee 1971). Curiosamente, muitas algas vermelhas possuem a característica de abrigar núcleos transmissivos, ou seja, núcleos que podem se mover de uma célula para outra por meio dessas conexões (Villarreal 2008). Esse também pode ser um fator importante no aumento da diversidade de retrotransposons em macroalgas vermelhas, uma vez que, por meio dessas estruturas, retroelementos de uma célula poderiam se deslocar para células adjacentes e também “infectá-las”, permitindo assim o aumento da diversidade de retrotransposons. No entanto evidências adicionais são necessárias para estabelecer uma conexão entre os *pit connection* e a diversidade retrotransposons em *P. umbilicalis*.

Foi possível observar em *P. umbilicalis* retroelementos com cromodomínio do tipo I e do tipo II. Até então, a única espécie vegetal que apresentou ambos os tipos foi *Selaginella moellendorffii* (Novikov et al. 2012). Além dos resíduos de aminoácidos conservados, os CHDs possuem um motivo característico, chamado de *chromo-box*, essencial para a formação de um núcleo hidrofóbico que pode estar relacionado com a integração do retroelemento em regiões de heterocromatina (Kordis 2005). Embora a função dos CHDs de retrotransposons

seja pouco conhecida, segundo proposto por Novikov et al. (2012), a evolução dos cromodomínios pode estar relacionado com a adaptação dos cromovírus às alterações na organização da cromatina das plantas. Além disso, a diferença observada no *chromo-box* de diferentes clados de cromovírus também pode estar relacionado com a especificidade de integração de cada um (Neumann et al. 2019).

Os CHDs de LTR-RTs encontrados em *Selaginella* são considerados um estágio de transição entre CHD tipo I e CHD tipo II. Curiosamente, o cromodomínio presente no Clado F, classificado como sendo do tipo II, apresenta o motivo da *chromo-box* similar ao encontrado nos elementos ditos de transição em *S. moellendorffii*. Levantando assim a hipótese de que a diferenciação do cromodomínio do tipo I para o tipo II possa ter ocorrido anteriormente a divisão das Rhodophytas e Plantas Verdes, e não durante a divisão das monocotiledôneas e dicotiledôneas, como se acreditava até então (Novikov et al. 2012). Não podemos descartar também a hipótese de que esse processo possa ter ocorrido em dois momentos distintos do processo evolutivo, no entanto a complexidade da evolução de retroelementos em plantas sem sementes e algas necessitam de mais estudos.

Apesar de uma grande quantidade de novos elementos ter sido descrita neste trabalho, observamos que nossa abordagem de identificação de elementos transponíveis pode fornecer apenas um limite inferior para o número de elementos *Ty3-Gypsy* de *P. umbilicalis*. Visto que o genoma em questão é representado por muitos scaffolds de pequenas sequências, ao invés de uma sequência por cromossomo. Isso pode impedir a identificação de alguns elementos que possam estar localizado no final da sequência de um scaffold. Além disso, os parâmetros estruturais utilizados em nossa busca foi baseado em sequências de retroelementos já conhecidas, impossibilitando assim que elementos com estruturas diferentes pudessem ser identificados.

## REFERÊNCIAS

BLOUIN, Nicolas et al. Seeding nets with neutral spores of the red alga *Porphyra umbilicalis* (L.) Kützinger for use in integrated multi-trophic aquaculture (IMTA). **Aquaculture**, v. 270, n. 1-4, p. 77-91, 2007.

BLOUIN, Nicolas A. et al. *Porphyra*: a marine crop shaped by stress. **Trends in plant science**, v. 16, n. 1, p. 29-37, 2011.

BRAWLEY, Susan H. et al. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). **Proceedings of the National Academy of Sciences**, v. 114, n. 31, p. E6361-E6370, 2017.

BREHM, Alexander et al. The many colours of chromodomains. **Bioessays**, v. 26, n. 2, p. 133-140, 2004.

CARARETO, Claudia Marcia Aparecida; MONTEIRO-VITORELLO, Claudia Barros; VAN SLUYS, Marie-Anne. **Elementos de transposição: diversidade, evolução, aplicações e impacto nos genomas dos seres vivos**. SciELO-Editora FIOCRUZ, 2015.

EICKBUSH, Thomas H.; JAMBURUTHUGODA, Varuni K. The diversity of retrotransposons and the properties of their reverse transcriptases. **Virus research**, v. 134, n. 1-2, p. 221-234, 2008.

EDGAR, Robert C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic acids research**, v. 32, n. 5, p. 1792-1797, 2004.

ELLINGHAUS, David; KURTZ, Stefan; WILLHOEFT, Ute. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. **BMC bioinformatics**, v. 9, n. 1, p. 18, 2008.

FELSENSTEIN, Joseph. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, v. 39, n. 4, p. 783-791, 1985.

FESCHOTTE, Cédric; JIANG, Ning; WESSLER, Susan R. Plant transposable elements: where genetics meets genomics. **Nature Reviews Genetics**, v. 3, n. 5, p. 329, 2002.

FINNEGAN, David J. Transposable elements. **Current opinion in genetics & development**, v. 2, n. 6, p. 861-867, 1992.

GANTT, Elisabeth et al. *Porphyra*: complex life histories in a harsh environment: *P. umbilicalis*, an intertidal red alga for genomic analysis. In: **Red algae in the genomic age**. Springer, Dordrecht, 2010. p. 129-148.

GAO, Xiang et al. Chromodomains direct integration of retrotransposons to heterochromatin. **Genome research**, v. 18, n. 3, p. 359-369, 2008.

GORINSEK, Benjamin; GUBENSEK, Franc; KORDIS, Dusan. Evolutionary genomics of chromoviruses in eukaryotes. **Molecular biology and evolution**, v. 21, n. 5, p. 781-798, 2004.

HAWKINS, Jennifer S. et al. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. **Proceedings of the National Academy of Sciences**, v. 106, n. 42, p. 17811-17816, 2009.

HEDGES, S. Blair. The origin and evolution of model organisms. **Nature Reviews Genetics**, v. 3, n. 11, p. 838, 2002.

HERPIN, Amaury et al. Transcriptional rewiring of the sex determining *dmrt1* gene duplicate by transposable elements. **PLoS genetics**, v. 6, n. 2, p. e1000844, 2010.

KAZAZIAN, Haig H. **Mobile DNA: finding treasure in junk**. FT Press, 2011.

KIDWELL, Margaret G.; LISCH, Damon R. Perspective: transposable elements, parasitic DNA, and genome evolution. **Evolution**, v. 55, n. 1, p. 1-24, 2001.

KHAN, Hameed et al. Retrotransposons and tandem repeat sequences in the nuclear genomes of cryptomonad algae. **Journal of molecular evolution**, v. 64, n. 2, p. 223-236, 2007.

KOHANY, Oleksiy et al. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. **BMC bioinformatics**, v. 7, n. 1, p. 474, 2006.

KOONIN, Eugene V.; ZHOU, Shubo; LUCCHESI, John C. The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. **Nucleic acids research**, v. 23, n. 21, p. 4229-4233, 1995.

KORDIŠ, Dušan. A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. **Gene**, v. 347, n. 2, p. 161-173, 2005.

LEE, Robert Edward. The pit connections of some lower red algae: ultrastructure and phylogenetic significance. **British Phycological Journal**, v. 6, n. 1, p. 29-38, 1971.

LETUNIC, Ivica; BORK, Peer. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. **Bioinformatics**, v. 23, n. 1, p. 127-128, 2006.

LINDAUER, Andreas et al. Reverse transcriptase families and a copia-like retrotransposon, *Osser*, in the green alga *Volvox carteri*. **FEBS letters**, v. 319, n. 3, p. 261-266, 1993.

LLORENS, Carlos et al. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology direct*, v. 4, n. 1, p. 41, 2009.

LLORENS, Carlos et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. **Nucleic acids research**, v. 39, n. suppl\_1, p. D70-D74, 2010.

MACARTAIN, Paul et al. Nutritional value of edible seaweeds. **Nutrition reviews**, v. 65, n. 12, p. 535-543, 2007.

MALIK, Harmit S.; EICKBUSH, Thomas H. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. **Journal of Virology**, v. 73, n. 6, p. 5186-5190, 1999.

MARCHLER-BAUER, Aron et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. **Nucleic acids research**, v. 45, n. D1, p. D200-D203, 2016.

MARÍN, Ignacio; LLORENS, Carlos. Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. **Molecular biology and evolution**, v. 17, n. 7, p. 1040-1049, 2000.

MCCLINTOCK, Barbara. Chromosome organization and genic expression. In: **Cold Spring Harbor symposia on quantitative biology**. Cold Spring Harbor Laboratory Press, 1951. p. 13-47.

NEUMANN, Pavel et al. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. **Mobile DNA**, v. 10, n. 1, p. 1, 2019.

NODA, Hiroyuki. Health benefits and nutritional properties of nori. **Journal of Applied Phycology**, v. 5, n. 2, p. 255-258, 1993.

NOVIKOV, Anton; SMYSHLYAEV, Georgiy; NOVIKOVA, Olga. Evolutionary history of LTR retrotransposon chromodomains in plants. **International journal of plant genomics**, v. 2012, 2012.

NOVIKOVA, Olga et al. Novel clades of chromodomain-containing Gypsy LTR retrotransposons from mosses (Bryophyta). *The Plant Journal*, v. 56, n. 4, p. 562-574, 2008.

NOVIKOVA, Olga; SMYSHLYAEV, Georgiy; BLINOV, Alexander. Evolutionary genomics revealed interkingdom distribution of Tcn1-like chromodomain-containing Gypsy LTR retrotransposons among fungi and plants. **BMC genomics**, v. 11, n. 1, p. 231, 2010.

NUNES, Renata C. et al. Structure and distribution of centromeric retrotransposons at diploid and allotetraploid *Coffea* centromeric and pericentromeric regions. **Frontiers in plant science**, v. 9, p. 175, 2018.

OHNO, Susumu. So much 'junk' DNA in our genome. In: **Evolution of Genetic Systems, Brookhaven Symp. Biol.** 1972. p. 366-370.

OKONECHNIKOV, Konstantin et al. Unipro UGENE: a unified bioinformatics toolkit. **Bioinformatics**, v. 28, n. 8, p. 1166-1167, 2012.

OU, Shujun; JIANG, Ning. LTR\_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. **Plant physiology**, v. 176, n. 2, p. 1410-1422, 2018.

PEDDIGARI, Suresh et al. Two different clades of copia-like retrotransposons in the red alga, *Porphyra yezoensis*. *Gene*, v. 424, n. 1-2, p. 153-158, 2008.



PÉREZ-ALEGRE, Mónica; DUBUS, Alain; FERNÁNDEZ, Emilio. REM1, a new type of long terminal repeat retrotransposon in *Chlamydomonas reinhardtii*. **Molecular and cellular biology**, v. 25, n. 23, p. 10628-10638, 2005.

PAZ, Rosalía Cristina et al. Diversity, distribution and dynamics of full-length Copia and Gypsy LTR retroelements in *Solanum lycopersicum*. **Genetica**, v. 145, n. 4-5, p. 417-430, 2017.

PRAY, Leslie A. Transposons: The jumping genes. **Nature education**, v. 1, n. 1, p. 204, 2008.

PUTEROVA, Janka et al. Satellite DNA and transposable elements in seabuckthorn (*Hippophae rhamnoides*), a dioecious plant with small Y and large X chromosomes. **Genome biology and evolution**, v. 9, n. 1, p. 197-212, 2017.

REBOLLO, Rita; ROMANISH, Mark T.; MAGER, Dixie L. Transposable elements: an abundant and natural source of regulatory sequences for host genes. **Annual review of genetics**, v. 46, p. 21-42, 2012.

RICE, Peter; LONGDEN, Ian; BLEASBY, Alan. EMBOSS: the European molecular biology open software suite. 2000.

SAITOU, Naruya; NEI, Masatoshi. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, v. 4, n. 4, p. 406-425, 1987.

SUTHERLAND, Judith E. et al. A new look at an ancient order: generic revision of the Bangiales (Rhodophyta) 1. **Journal of phycology**, v. 47, n. 5, p. 1131-1151, 2011.

THOMPSON, Julie D.; HIGGINS, Desmond G.; GIBSON, Toby J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic acids research**, v. 22, n. 22, p. 4673-4680, 1994.

VILLARREAL, Luis P. **Origin of group identity: viruses, addiction and cooperation**. Springer Science & Business Media, 2008.

VOLFF, Jean-Nicolas. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. **Bioessays**, v. 28, n. 9, p. 913-922, 2006.

XU, Zhao; WANG, Hao. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. **Nucleic acids research**, v. 35, n. suppl\_2, p. W265-W268, 2007.

ZHANG, Jianbo et al. Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. **Genes & development**, v. 23, n. 6, p. 755-765, 2009.